DOCUMENT RESUME

ED 090 299 TM 003 584

AUTHOR	Sewell, Alan F.; Dornseif, Allan W.
TITLE	Controlled Multivariate Evaluation of Open and Traditional Education at the Junior High School Level: The First-Year Report.
SPONS AGENCY	Spencer Foundation, Chicago, Ill.
PUB DATE	Apr 74
NOTE	31p.; Paper presented at the American Educational Research Association Annual Meeting (Chicago, Illinois, April 15-19, 1974)
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	*Academic Achievement; Evaluation Methods; Junior High Schools; Open Education; *Program Evaluation; Social Adjustment; Student Attitudes; *Traditional Schools
IDENTIFIERS	Matteson Illinois School District

ABSTRACT

Measures of attitude, academic achievement, and social and personality variables were employed in a year-long evaluation of progress and outcomes of open education and traditional education groups. Data were analyzed by multivariate analysis of variance and discriminant function techniques, using means of teacher ratings as the "success" criterion. The two groups were found to differ in a number of ways. Results are given, and the relative advantages and disadvantages of this evaluative methodology, based on random assignment of students, are discussed. (Author)

25.05

ED 090299

TM 003 584

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Controlled Multivariate Evaluation of Open and Traditional Education at the Junior High School Level: The First-Year Report

Alan F. Sewell
Evanston, Illinois

and

Allan W. Dornseif
Assistant Superintendent
School District 162
Matteson, Illinois

This study was made possible by a grant from The Spencer Foundation. The authors wish to acknowledge the assistance and cooperation of J. Schlenker, Jr., Superintendent, School District 162, and D. R. Stevenson, Principal, O. W. Huth Upper Grade Center. Computer service was provided by the Computer Center, University of Illinois Chicago Circle. Special thanks are due to W. R. Jarosz and N. D. Dyhdalo, DePaul University Computation Laboratory for their aid in keypunching and data processing; K. Billingham, DePaul University, for assistance in data collection; and M. J. K. Brown, DePaul University, for advice in design and analysis.

Controlled Multivariate Evaluation of Open and Traditional Education at the Junior High School Level: The First-Year Report

Alan F. Sewell and Allan W. Dornseif

Although undoubtedly some enterprising scholar could trace the ancestry of open education to the Greek akademeia (nevertheless, open education seems to be more socratic than platonic), the more remote of open education's contemporary ancestors was the educational philosophy of progressivism of some four or five decades ago. The immediate parents of the open education movement (as it has come to be described), however, were the philosophies and experiences of the informal classrooms of England, which emphasized the priority of student choice in the selection of learning activities and the student's conjoint role as learner-teacher.

While in theory open education suggests a movement toward certain educational objectives, in practice it has been most distinctive as a movement away from traditional educational methods. Growing criticisms of and disenchantment with much of traditional education, both in England and in the United States, has led to a rather wide range of educational innovations, the communality of which is their departure from traditional methods.

Awareness of the English experiments grew slowly during the early years of the last decade, but by the middle years of that decade, a considerable interest had developed in their possible applications to American education. Hawkins (1965, 1967), for example, attempted to identify significant elements of open education in order to provide a theoretical base for an American version, but his primary focus was science education. The momentum of such interest was substantially increased by the official sanction given the open education movement in England with the publication of the Plowden Report (1966) and by publi-

cation in the United States of a series of germinal articles and a book by Featherstone (1968, 1971a, 1971b, 1971c). In these publications Featherstone not only explored the English principles and practices in detail, but he also provided a number of thoughtful suggestions bearing on the potential implementation in American education.

Still further impetus was provided by Silberman's Crisis in the Classroom (1970), which recommended elements of open education a promising alternatives to traditional education in the United States. Similarly influential in establishing philosophical and conceptual bases for American models were The ESS Reader (1970) and Open Education: The Informal Classroom (Rathbone, 1971).

Possibly the earliest widespread application of concepts derived from the English informal classroom experiences took the form of the "free" schools which seemed to bloom everywhere in the United States toward the end of the last decade. These schools were -- and are -- characterized by a general laissez-faire approach to planning, instead emphasizing freedom, spontaneity, and individual choice. Many pointedly identified themselves as "alternative" schools, objecting to both the systematization and the methodologies of traditional education. In effect these schools deny the possibility of innovation and improvement within existing educational systems.

Perhaps in response to such challenges, the past few years have seen an enormous development of conceptualization and implementation of various forms of open education in the United States. Most of this development has occurred within the educational "establishment" and appears generally to be a compromise between the extremes of informality advocated by "free" schools proponents and the fairly rigid structure of traditional education. Typifying this compromise is Howard's (1968) version of open education, the "Developmental Classroom," which holds that a free or open approach to education still requires a comforting sense of structure.

Official sanction in the United States was gained when late in the last de-

cade North Dakota adopted open education as a systematic approach and, moreover, reorganized teacher training to ensure perpetuation of the adopted model. Similar acceptance has been granted or is being considered in several other States, including New Jersey, New York, and Vermont. Massachusetts' Educational Development Center has helped to organize a large number of open classrooms throughout the United States, following the basic pattern of the English informal schools. The Center also has developed materials and a useful bibliography (Barth & Rathbone, 1971) to aid in the planning and implementation of open education programs. Educational Testing Service has also published (1970) a useful guide to establishment and evaluation of open education programs.

How Open Is Open?

Because open schools and open classrooms, as distinct from "free" schools, have typically been introduced within existing educational systems in the United States, compromises, as noted earlier, have been effected between the concepts and methods of "pure" open education and the concepts and methods of traditional education. As a result there now exist hundreds of open classrooms and open schools which vary enormously. The open education movement has clearly not progressed sufficiently to have achieved any degree of standardization; indeed, it well may be that such standardization would prove to be the nemesis of open education. Nevertheless, until some consensus on openness is attained, evaluation of the fundamental worth of this approach must necessarily proceed piecemeal.

In recognition of this difficulty a number of attempts have been made to develop objective criteria of open education. Two of these attempts are particularly noteworthy. Walberg and Thomas (1971) studied the writings of leading British and American theorists of open education and, through content analysis, extracted eight basic or characteristic "themes." Subsequently they provided (1972) an instrument intended to objectify the degree of openness in an

educational environment. These characteristics or "themes" are: (1) provisioning for learning; (2) humaneness, respect, openness, warmth; (3) diagnosis of learning events; (4) instruction, guidance, extension of learning; (5) evaluation of diagnostic information; (6) seeking opportunities for professional growth; (7) self-perception of the teacher; (8) assumptions about children and the learning process. Significantly, 25 of the 50 items of the questionnaire-observation schedule concern the "provisioning" theme, suggesting a preponderant emphasis upon variables of the physical environment.

Traub et al. (1973) also used the theoretical literature of open education in an effort to identify critical characteristics. A 29-item rating scale, known as Dimensions of Schooling (DISC), was developed for use by teachers in reporting aspects of school or classroom life. These items were chosen to reflect ten characteristics of open education. A subsequent factor analysis of 25 of the 29 items produced six clusters of items: (1) individualization of instruction, (2) student independence, (3) environmental flexibility, (4) nongradedness, (5) flexibility of student evaluation, (6) flexibility of curricular materials. The emphasis in the definition of open education implied by the DISC appears to be upon individualization, a characteristic which is, again by implication, somewhat less directly important in the Walberg-Thomas definition. That is, scores on these two instruments might well be poorly correlated, suggesting that rather different criteria are involved; such a cross-validation study is now being conducted by the present authors.

Clearly some obscurity of distinction between openness as an environmental characteristic and openness as a program characteristic exists, justifying the classification by Traub et al. (1973) of "open space" schools and "open program" schools. The primary characteristic of the former is the architectural plan of the learning space and the kinds of learning and teaching behaviors necessitated and facilitated by absences of walls and fixed equipment. The primary characteristic of the latter appears to be some degree of individualization

of instruction.

Somewhat related to this characteristic of open program schools is the Individually Guided Education/Multi-Unit School (IGE/MUS) approach developed in Wisconsin. More an attempt to increase individualization and team teaching, this approach nonetheless contains elements of a philosophy of open education: emphasis upon spatial arrangements, deployment of teachers, and self-guidance by the student. The degree of systematization implicit in this approach reflects, however, the same kind of compromise noted earlier as generally applied in American models of open education. Also related in some degree to the philosophy of open education is the non-graded school; as noted previously, non-gradedness is one of the distinctive characteristics of open education in the view of Traub et al. (1973).

Currently there is no single "American model" of open education. Instead there are many models, each differing demonstrably from the others. They share in attempting to provide alternatives to traditional education, and very likely they constitute simultaneous evolutionary experiments in educational reform. But the evolutionary struggle between traditional and innovative education, and the struggle between competitive varieties of educational innovation will be settled less by biological than by social processes of natural selection, and in particular by the acceptances and preferences of the education-consuming public who support educational institutions and systems. Hopefully evaluative research will enable that public to make informed choices.

Some Relevant Research

Because of the relative recency of open education in America and because of the plethora of somewhat dissimilar models, evaluation of open education is still fairly primitive. Indeed, an appropriate evaluation model has not yet been developed. Most studies which purport to evaluate open education are conceptually or methodologically imperfect, most commonly because of one or more of the fol-

lowing limitations: (1) lack of control groups pursuing traditional programs; (2) non-random assignment of students to open groups and control groups; (3) use of groups at different schools, with a consequent confounding of socioeconomic variables; (4) insufficient definition of "openness" as an independent variable; (5) insufficient rationale for selection of dependent variables; (6) use of inappropriate analytical designs and techniques. Intrusion of such flaws into the design, conduct, or analysis of results of these studies clearly prevents generalization of results and frequently justifies scepticism concerning results.

At the present stage of evaluation strategies applicable to open education, it seems more appropriate to focus upon the nature of expected differences rather than upon the results of specific studies. Various studies have sought differences in academic achievement, in aspects of personality, in social behavior, and in self-concept. Burnham (1971), for example, compared intelligence, achievement, and behavioral differences of first-graders in open and traditional schools. Sackett (1971) measured IQ, achievement, and self-concept of sixth-graders in three different educational settings. Killough (1971) anticipated interaction between sex and achievement as a function of open education. Warner (1971) compared open and traditional fourth grade classes in achievement, grouping time, material usage, and classroom climate. Wilson et al. (1972) focused on measures of curiosity and productive thinking of students in open plan schools. Kennedy and Say (1971) compared second- through fifth-grade students of open and traditional schools in academic achievement. Carbonari (1971) studied differences in personality factors as a function of length of experience in an open school. The results of these studies are typically inconclusive and/or inconsistent with other studies. And because "openness" is typically unquantified, no generalizability can be derived from them.

Hopefully, with the development of instruments to measure "openness" or to establish the degree of difference between open and traditional programs, a major

limitation of such studies will be eliminated. One such study has already appeared and, within its own limitations, can well serve as a model for evaluative trends in open education. Having developed a means of quantifying openness, Traub et al. (1973) proceeded to apply that means in an evaluation of student outcomes as a function of degree of openness. A number of Canadian elementary schools were classified according to architectural openness ("open architecture," "mixed architecture," or "closed architecture"), degree of program openness on the basis of DISC ratings, and character of student population (based on number of students employing English as a second language, hence, presumably representing cultural homogeneity or heterogeneity). In terms of cognitive skills, no consistent pattern of advantages could be claimed for any given combination of architectural openness and program openness, at least for those schools in which the vast majority of students did not come from homes using English as a second language. However, in those schools in which more than 30% of the students were from homes using English as a second language (whom the authors seem to find similar to "so-called inner city kids"), a consistent pattern of cognitive superiority was found in traditional programs. Results of non-cognitive measures again showed no clearly consistent pattern, although the investigators report that many of the differences between groups are in the directions which would be predicted by open education proponents, and that these differences are more clear-cut and more striking for older students (11-year-olds) than for younger students (8-year-olds).

Traub et al. acknowledge several of the deficiencies of their study, yet it quite likely represents the most careful and conscientious research to date in the area of open education. Of particular merit are the rather large number of schools involved, the classification of degrees and kinds of openness, and the number of variety of dependent variables considered. Nevertheless, the results are more provocative than definitive, even within the limitations of the population studied.

11

Although the English experiences in open education and large-scale adoption of American variants seems to imply an established superiority of open education over traditional education, the typically equivocal results of recent research recommend a more cautious approach. There appears to be no urgency to a belief that "openness" is necessarily desirable for all types of students, nor for all ages of students, nor for all educational levels. It well may be, as results of the study by Traub et al. (1973) suggest, that for certain students traditionally structured education yields more satisfactory outcomes. With such a possibility in view, the present study was undertaken.

Origin and Operation of the OSCAR Project

This evaluation project was conducted at the O. W. Huth Upper Grade Center, School District 162, Matteson, Illinois. When it became necessary to construct an addition to the existing Upper Grade Center building, a decision was reached to implement an experimental open education program which would include seventh and eighth grade boys and girls in a single large classroom and which would be staffed by an interdisciplinary team of teachers.

The District, located in south suburban Cook County, includes all or parts of five communities which encompass quite a wide socioeconomic range. Approximately 3,400 children attend the District's eight schools. Of these, some 2,600 children attend seven primary schools, and about 800 attend the District's single junior high school (the Upper Grade Center). Children attend local schools through sixth grade and then complete seventh and eighth grade at the Upper Grade Center. Hence the District's heterogeneity is fully expressed only at the junior high school level.

The open classroom of the Upper Grade Center was planned and designed as part of a total remodeling and enlargement program. A space equivalent to that of five traditional classrooms and their adjacent hallway was allocated to the open classroom: a single, undivided area approximately 140 feet long and 40 feet

wide. The fortuitous acronym OSCAR was soon applied to the open classroom project, memorializing a former District superintendent and representing "Open Space for Conceptualizing Attitudes and Responsibilities."

The OSCAR program, as developed, called for staffing by four teachers and two teacher-aides for a total of 140 seventh- and eighth-graders. One teacher was chosen from each of four major discipline areas: language arts, mathematics, science, and social studies. These teachers were chosen on the basis of expressed interest in and probable capability for the OSCAR program. During the summer preceding inauguration of the program, they attended a ten-day multitalent development training program which emphasized the discovery and development of talents unique to each student. The OSCAR concept required these teachers to plan and work together in content areas and to develop a multidisciplinary approach to subject matter.

Initiation of the open classroom project was planned for the 1972-73 school year. In the Spring of 1972 two groups of children were randomly chosen: an OSCAR group and a Control group, with 140 children in each group. Selection of the control group was based upon an intention to provide clear and continuing evaluation of the open classroom project. Half of each group was chosen from the Upper Grade Center's existing enrollment of then-seventh-graders, and half was chosen from the District's population of then-sixth-graders (who would attend the Upper Grade Center as seventh-graders during the 1972-73 school year). Each group included 35 seventh grade boys, 35 seventh grade girls, 35 eighth grade boys, and 35 eighth grade girls. The random assignment procedures were designed to ensure proportional representation of the entire District on the basis of each primary school's contribution to the junior high school population.

OSCAR students were informed of their assignment during Spring, 1972. Control group students were never specifically informed of their function but were identified at times of assembly for testing as the principal's "Special Group" in an attempt to control for Hawthorne effect.

Early months of the OSCAR program were somewhat confused, primarily as a result of construction delays and the newness of the program. The OSCAR teachers found it necessary to modify their plans with respect to multi-talent development and multidisciplinary planning. The school day was instead organized around the four major disciplines in the morning, with the amount of time devoted to each discipline determined cooperatively by the teachers. Projects and modified contract assignments were the basic activities of students in all subjects except mathematics, wherein an individualized skill development materials kit was used. During the afternoon, OSCAR students were scheduled into other school programs, such as physical education, home economics, industrial arts, music, and art. While most of the OSCAR group was out of the open classroom during this time, about 25% of the students remained in the room to complete individual assignments, work in small study groups, or receive individual or small group tutoring.

Control group students, meanwhile, pursued the "traditional" departmentalized curriculum and schedules of the Upper Grade Center, changing rooms and classmates at the end of each class period. These students were never together as a group except when assembled for testing purposes associated with this study. No attempt was made to encourage interdisciplinary teaching or the introduction of any unusual teaching techniques among non-OSCAR teachers. For purposes of this study, however, four non-OSCAR teachers, one in each of the four primary discipline areas, were selected as "control" teachers without their knowledge; this was done to permit some comparisons of teacher behaviors.

While no attempt was made during the course of the study to evaluate the relative "openness" of the OSCAR program -- as, for example, by the DISC (Traub et al. (1973) or by the Walberg and Thomas (1971) scale -- it is clear that the OSCAR program was designed to implement a fairly limited degree of "openness" as compared to the informality, freedom, and structurelessness of "free" schools. Nevertheless the freedom afforded the OSCAR students was substantially greater

than that accorded the control students. The OSCAR program was intended to incorporate as many of the general characteristics of "open education" as seemed appropriate to the needs and abilities of students, teachers, and parents.

This evaluation study was primarily focused upon student outcomes and was intended to determine (1) whether the open and traditional programs produced overall differences in these outcomes, and (2) whether student outcomes differed as a function of student characteristics, such that criteria for assignment of individual students to a given program could be formulated.

Method

The basic design strategy of this evaluation project was to establish a research model which would permit continuing, definitive comparisons in many outcome areas. The first-year plan, then, amounted to a "shotgun" approach to evaluation, hopefully to provide direction for subsequent, more specific evaluative efforts within an established model.

A variety of measures, in the general areas of academic, personal, and social development, were used. The Stanford Achievement battery, long employed in this District, was used to measure academic achievement in the areas of paragraph meaning, spelling, language, arithmetic computation, arithmetic concepts, arithmetic application, social studies, and science. Baseline data for the eighth-graders were available from the final testing of the previous academic year, but the seventh-graders had (as sixth-graders) completed a different form of the battery. The entire battery was administered twice during the evaluation year: at midpoint and at the end of the year.

Various aspects of personal and social adjustment were measured by the Bell Adjustment Inventory. This instrument, although standardized for high school students, was chosen because of its reported reliability and its wide use among student populations. The Inventory was administered three times: near the beginning of the year, at midpoint, and near the end of the year. Data from the

first administration, however, were discarded because (1) through administrative error a large number of the control subjects did not complete the Inventory, and (2) a subsequent study of the comprehensibility of the items by the subjects led to the elimination of approximately 25 items which were not understood by the subjects. To prevent difficulties occasioned by the relatively advanced reading level of the Inventory, all items were tape-recorded for the second and third administrations; subjects were not given the Inventory booklets.

Other instruments were administered only once, near the end of the school year. These instruments were the California Psychological Inventory (Gough, 1957, revised edition 1969), which provides data in 18 areas of personality characteristics; the Piers-Harris Children's Self Concept Scale, which provides a single measure of self-concept; the School Inventory (Bell, 1963), which measures attitude toward school; and a measure of "locus of control" (Rotter, 1966) to provide data on the internal-external orientation of the student.

On three occasions during the year, the four OSCAR and four "control" teachers were observed by a trained observer who recorded behaviors according to Flanders' (1966) categorical analysis of classroom interactions. Finally, at the end of the year the four OSCAR teachers were asked to rate each OSCAR student on a five-point scale on each of four dimensions of performance: attitude, knowledge, skills, and sociability. The following behavioral definitions were provided to each teacher:

> Attitude: Student displays positive attitudes toward school, teachers, other school personnel, and other students.
>
> Knowledge: Student demonstrates mastery of academic content appropriate to his/her age, grade level, and apparent ability.
>
> Skill: Student demonstrates application of academic content within school and displays ability to apply academic content in non-academic settings.
>
> Sociability: Student demonstrates respect for the rights and

feelings of others and demonstrates ability to work effectively and cooperatively with others.

Teachers were instructed to evaluate each student independently of other students and to use each rating dimension independently of the others.

Lists of the control group students were provided to all teachers in the school's departmental program, and each teacher was asked to rate each student who had been in his classes during the year. This procedure was necessary for the control group since the four "control" teachers had not instructed all the control group students during the year.

Results

Two varieties of data analysis were conducted: absolute analyses, which focussed upon whether any absolute differences in outcome measures could be demonstrated between the OSCAR and Control groups; and relative analyses, the object of which was to detect differences in the nature of program outcomes. For the first of these varieties, the primary independent variable was the grouping variable (OSCAR vs. Control); independent variables of secondary interest were grade level (seventh vs. eighth), and sex (male vs. female). For the second variety of analysis, the single independent variable was program outcome as either "success" or "failure."

Since three major evaluative instruments employed in this study (Stanford Achievement battery, Bell Adjustment Inventory, California Psychological Inventory) consisted of numbers of intercorrelated subscales, data from these instruments were analyzed through multivariate analysis of variance techniques. Where overall significance was demonstrated, post-hoc univariate *F*s and step-down *F*s were computed. Data from single-scale instruments were analyzed through appropriate univariate analysis of variance models. In all cases analyses were based upon raw scores rather than grade-level equivalents, standardized scores, or the like.

Absolute Analyses

For purposes of economy, results will be summarized for each instrument. Unless otherwise noted, scales not mentioned did not significantly differentiate the two groups.

Stanford Achievement Battery. The multivariate F for Group was significant ($F = 6.009$, $df = 8/201$, $p < .001$). Two significant subscale differences were found by post-hoc analyses: Social Studies ($F = 3.729$, $df = 1/208$, $p < .01$) and Science ($F = 14.844$, $df = 1/208$, $p < .001$). In both cases the OSCAR group outperformed the control group. The multivariate Fs for Grade Level and for Sex were also significant; as would be expected, eighth-graders outperformed seventh-graders, and females outperformed males in those areas in which differences occurred.

The multivariate F for the Group x Grade Level interaction was also significant ($F = 2.718$, $df = 8/201$, $p < .01$). Post-hoc analyses showed that in Spelling and in Language, the seventh grade OSCAR students outperformed other subgroups, with the eighth grade Control students second-best. In Arithmetic Computation and Arithmetic Concepts, the eighth grade Control students excelled, with the seventh grade OSCAR students second-best. No other interactions were significant.

Bell Adjustment Inventory. The multivariate F for Group was significant ($F = 3.843$, $df = 6/206$, $p < .01$). Post-hoc analyses showed that significant differences occurred on the Submissiveness-Self-Assertion scale ($F = 6.417$, $df = 1/206$, $p < .01$) and the Masculinity-Femininity scale ($F = 11.079$, $df = 1/206$, $p < .001$); in both cases higher scores were achieved by the Control group students.

The multivariate F for Grade was significant ($F = 2.310$, $df = 6/206$, $p < .05$). The only scale contributing significantly to this difference was the Submissiveness-Self-Assertion scale ($F = 4.515$, $df = 1/206$, $p < .05$); here the seventh-graders achieved higher scores than the eighth-graders.

As would be expected, the multivariate F for Sex was also significant (F = 39.741, df = 6/206, p <.001). Post-hoc analyses showed females to score higher in Emotionality than males, and males to score higher than females in Masculinity-Femininity. None of the interactions was significant.

California Psychological Inventory. Of the 18 scales of the Inventory, only one differentiated the two groups: Sense of Well-Being (univariate F = 5.010, df = 1/165, p <.05). The Control mean for this scale was higher than the OSCAR mean.

Piers-Harris Children's Self Concept Scale. This scale provides a single score; hence results were analyzed by ANOVA. The F for Group was the only significant finding (F = 7.215, df = 1/208, p <.01). The Control mean was higher than the OSCAR mean.

The School Inventory. This attitude inventory yields a single score; hence data were analyzed by ANOVA. None of the effects attained statistical significance.

Locus of Control. Here, too, a single score is produced, and ANOVA was employed. None of the effects was significant.

Relative Analyses

The primary objective of the several relative analyses conducted was to permit differential characterization of the two educational programs and to identify student characteristics appropriate to each program. The paucity of significant findings through absolute analyses was, conceivably, due to the presence of students in each group who might have functioned significantly differently in the other group. Hence the focus of interest in the relative analyses was the student's relative performance within his group.

Teachers' ratings of OSCAR and Control students were chosen as criteria for the relative analyses. In the case of the OSCAR teachers, each teacher's four-dimension rating of students was converted into a z distribution for that teacher,

and on the basis of this distribution a mean z rating across the four areas was obtained for each student. An overall mean z rating for each student was derived from the mean z rating provided by each teacher; this final rating, then, was the end-product of 16 separate ratings by the four OSCAR teachers, corrected by the z transformation for differences in rating styles. A similar procedure was followed for Control group students, except that in some cases fewer than four teachers rated a student and, in consequence, each such student was dropped from the analysis; and in some cases more than four teachers rated a student, but only the "first" four ratings were used.

On the basis of these mean z ratings, each group was then divided as near the median as possible; those above the median were classified as "successful" students, while those below the median were classified as "unsuccessful" students. These admittedly arbitrary classifications permitted analyses of the "success-failure" characteristics of each group and, hence, each program. Only students who completed all measures employed in this study were included in these analyses; the totals thus available were insufficient for analysis by sex and grade level. Hence Group was the only independent variable employed in these analyses.

In the first analysis "successful" students were compared to "unsuccessful" students, regardless of group. Subsequent analyses compared various subgroupings of "successful" and "unsuccessful" students. Each analysis was a one-way multivariate analysis of variance, followed, where indicated, by univariate F tests. The results of these analyses are shown in Table 1.

- -

Insert Table 1 about here

- -

Teacher-Related Measures

The study provided two types of measures of teacher characteristics. Flan-

ders analyses of observations of teachers' behavior patterns in classroom interactions were conducted on three separate occasions. None of these analyses showed significant differences in patterns displayed by OSCAR teachers and by Control teachers. Review of the raw data showed that very substantial between-teachers variance occurred in both groups, and, with so small a sample size (four teachers in each group), very dramatic differences would have been required in order to attain statistical significance.

A second approach to the analysis of teachers' behaviors and styles was suggested by the student ratings provided by each teacher. Because each of a number of teachers had rated a relatively few students in the Control group, comparisons of rating characteristics were not possible. However, each of the four OSCAR teachers had rated the same 131 OSCAR students, and it was thereby possible to compare ratings. Intercorrelations of the teachers' mean z rating for each student are shown in Table 2.

- -

Insert Table 2 about here

- -

Although grade level and sex were not included as independent variables in the relative analyses of program outcomes, it was possible to analyze the contributions of these factors to the OSCAR and Control teachers' ratings. Tests of differences in mean z ratings of seventh and eighth grade students in the "success" and "failure" subgroups were conducted. For the OSCAR "success" subgroup, the difference was significant ($t = 7.670$, $df = 66$, $p < .001$); seventh and eighth grade OSCAR "unsuccessful" students did not differ significantly in mean ratings. The point-biserial correlation coefficient between grade level and "success" in the OSCAR group was .213 ($p < .01$). Grade level was not found to be significantly related to mean z rating in the Control group.

The relation of sex to "success" and "failure" in each group was tested by a chi-square test of independence. Within the OSCAR group these two variables

were found to be significantly related ($\chi^2 = 8.98$, $df = 1$, $p < .01$). The Goodman-Kruskal index of predictive association (Hayes, 1963, pp. 608-609) was computed as .183; that is, knowing the sex of the student reduces the probability of error in predicting success-failure by 18.3%. A point-biserial correlation coefficient between sex and mean z rating was also calculated for the OSCAR students; the obtained coefficient was .407 ($p < .001$), showing good agreement with the Goodman-Kruskal index. The mean difference between male and female z ratings is not, however, statistically significant. The chi-square test between sex and success-failure within the Control group showed that in this group the two variables are independent ($\chi^2 = 1.14$, $df = 1$, $p > .05$).

Discussion

Educational innovation, like any other kind of innovation, must justify itself on either or both of two bases: (1) that it accomplishes traditional objectives more efficiently; (2) that it accomplishes new (non-traditional) and desirable objectives. Hence the accomplishments of the OSCAR program must be evaluated finally in terms of those two bases.

Traditional Objectives

Mastery of academic content is probably the most "traditional" of educational objectives. Laypersons and probably most professional educators would agree that the fundamental objective of education is to "learn the three R's." At the high school and junior high school levels the traditional emphasis upon academic achievement is demonstrated by departmental organization, with each content area staffed by teachers possessing special subject-matter competences. It is quite likely that any educational program which demonstrated decrements in academic achievement would rapidly be rejected by public and educators alike.

Fortunately, the OSCAR program has demonstrated no such decrements; indeed, in two of four major subject-matter areas (social studies and science) the open

education program appears to have demonstrated superior results. Performance of students in both groups exceeded national norms to some degree; mean grade level equivalents over the eight areas evaluated by the Stanford Achievement battery were as follows: seventh grade OSCAR 8.14, seventh grade Control 7.22, eighth grade OSCAR 8.49, eighth grade Control 8.35. Hence the superior academic performances of the OSCAR students cannot be attributed to any obvious deficiencies in the traditional program.

The interactive effects of grade level and type of program upon academic achievement tempt a conclusion that open education programs may be better suited to younger students. Other findings of this study, however, reduce the urgency of such a conclusion. In any event, it can quite safely be concluded that the open education program of this school district is quite compatible with traditional educational objectives and may, in fact, promote a more efficient accomplishment of those objectives.

Non-Traditional Objectives

In terms of traditional objectives, success in school can be measured as a direct function of academic achievement. But any program which proclaims itself "alternative" or "innovative" must, by definition, imply a different kind of success. Although the philosophies and theories of traditional education generally agree that the objectives of education should extend beyond the purely cognitive realm, the theoretical literature of open education much more strongly emphasizes non-cognitive objectives. A strong implication exists that "success" in open education includes the student's performance and development in attitudinal and social areas as well. Since these accomplishments are not necessarily attained at the expense of cognitive skills, open education implies a wider, more encompassing definition of "success."

To be sure, a definition of "success in school" is not readily achieved. Prior to adoption of the teacher rating strategy employed in this study, the

four open classroom teachers and the four "control" teachers were asked to provide lists of behavioral objectives which governed their teaching. The lists provided by the two groups of teachers were essentially indistinguishable. While neither list was particularly noteworthy as an example of behavioral objectives, both included items which clearly extended beyond the purely cognitive realm.

The rationale of the teacher rating strategy held that "success" within a given program could, perhaps, best be defined by those student characteristics which significantly differentiated "successful" from "unsuccessful" students. That is, the program's (or teacher's) objectives could be inferred from the differential characteristics of those students who did and who did not meet those objectives.

The patterns of significant differences shown in Table 1 make it clear that "success" is not differently defined by open classroom teachers and traditional teachers: in both the major emphasis is upon academic achievement. The student who is "successful" in the traditional program is likely to be equally "successful" in the OSCAR program. And there is no reason to believe that the student who is "unsuccessful" in the OSCAR program will attain "success" in the traditional program. The emphasis upon academic achievement in both programs is probably also reflected in a common personality characteristic of the "successful" student: responsibility. The lack of difference in student outcomes in the two programs supports a belief that the objectives of the two programs do not differ; whether intentionally or not, the open classroom teachers have been pursuing conventional, traditional objectives.

Program Homogeneity

No undertaking, whatever its nature, can be expected to reach its objectives unless its participants agree on those objectives. Hence it is important to consider the degree to which agreement on program objectives has been achieved. Again the teachers' ratings of students are of use: agreement by the teachers as

to the relative "success" of each student, reflected in the intercorrelations of ratings shown in Table 2, can serve as an indicator of inter-teacher agreement on program objectives and, thus, of the homogeneity of the program.

The intercorrelations of Table 2 provide a somewhat ambiguous picture. While most of the intercorrelations are uncomfortably low, ratings by two of the four teachers are quite highly correlated. Several possible explanations are available: (1) these correlations actually are conditioned by the similarities and dissimilarities of the subject-matters areas themselves; (2) the correlations reflect fairly broad academic abilities of the students, in the sense that a student who does well in one area is likely to do well in other areas; (3) the correlations do actually reflect teacher communalities in objectives and methods. The sheer magnitude of the higher correlations of Table 2 tend to make the first two of these explanations untenable; the correlations substantially exceed the magnitude of correlations between Stanford subtests in the present study. Thus it appears that to a considerable degree at least two of the four teachers have been able to reach agreement on objectives, whether or not that agreement has ever been made explicit.

Unfortunately it was not possible to calculate similar intercorrelations for the control students' teachers, due to the relatively small number of students rated by each of the school's departmental teachers. Inspection of the ratings, however, suggested wide variability and rather poor agreement in ratings. This lends some credence to a belief that inter-departmental programs generally display rather poor homogeneity.

Special Problems of the Open Classroom

One of the original purposes of this evaluation study was to derive differential profiles of "successful" OSCAR students and "successful" traditional program students. Results of the study, however, demonstrate the futility of such an attempt at the present time: successful students in the two programs do

not differ because the programs do not, as yet, differ. Certain of the findings, however, paradoxically lend optimism to an expectation that special problems of the OSCAR program will require subsequent program differentiations and establishment of differential program objectives.

As noted previously, "success" in both the open classroom and the traditional classroom depends primarily upon academic achievement. In the open classroom but not in the traditional classroom, "success" is also positively correlated with sex (females having the advantage) and with grade level (seventh-graders having the advantage). "Successful" open classroom students also are characterized by "communality," which is essentially defined as a sense of belonging to a group and pursuing group objectives; a relative absence of this characteristic is found in "unsuccessful" open classroom students, but "communality" does not appear to be related to success in the traditional program. The "successful" OSCAR student also differs from his counterpart in the traditional program in having a lesser "sense of well-being" and a significantly lower self-concept. And, finally, "successful" OSCAR students are more internally controlled than "unsuccessful" students in this group.

Taken together, these findings present a fairly consistent portrait of the "successful" OSCAR student and, hence, of desirable student characteristics and behaviors within the open classroom. The student "most likely to succeed" in the OSCAR program's first year was a hard-working, self-disciplined, achievement-oriented seventh grade girl; she cooperates with her teachers, yet has a sense of personal dissatisfaction.

The data-gathering processes of this study were occasionally supplemented by interviews with teachers and students of the OSCAR program. The results of these interviews tend to support the general validity of the preceding description of the "successful" student. Teachers' comments typically focussed upon problems of classroom management, discipline, order, and noise; individual students most easily recalled were usually "problem" students, and the problems

were usually disciplinary in nature. Students, too, commented upon classroom noise and an atmosphere of apparent confusion. Indeed, as the year-end neared, a number of seventh-graders petitioned the school's principal for transfer to the traditional program for the following year. While inevitably some of the reasons for these requests concerned feelings about individual teachers, a large number of the given reasons dealt with the student's difficulties in adjusting to the atmosphere of the open classroom and its consequent demands upon self-discipline. Many of the students remarked, for example, that they were unable to complete assignments unless, as more commonly occurs in the traditional program, an immediate due-date is specified. Interviews also indicated that to some extent a morale problem has been generated within the OSCAR program through negative comments by non-OSCAR teachers; some of these teachers, apparently unconvinced of the program's worth and envious of the space and facilities allocated to it, had suggested that the "special" quality of the OSCAR program was that it was remedial!

Thus it appears that this open education program and its installation may have presented challenges to the entire school population, students, teachers, and administrators: challenges which were unexpected and with which the school was unprepared to cope. While the space allocated to the OSCAR program, and the number of children assigned to that space, approximates the school-wide average, the space itself demands new and different student and teacher behaviors. Noise and a degree of apparent confusion are inevitable consequences of housing so large a group of students in so large a space, and the teacher's probable first concern is with attaining order and discipline. This concern seems clearly reflected in the personal characteristics of the "successful" OSCAR student. It is to be expected that as their experience accumulates the OSCAR teachers will develop coping strategies which will enable them to focus more effectively upon their students' non-cognitive development; that is, "success" will be redefined to expand concepts of responsibility and sociability and to de-emphasize concepts

of order and discipline. The first year of this program has clearly demonstrated that academic standards can be maintained. With the teachers' increasing confidence and competence, the potential contributions of open education to the student's personal and social development can more meaningfully be assessed.

The findings of this study suggest some general observations pertinent not only to this particular open education installation but probably equally relevant to other present and contemplated programs:

1. Clearly a critical requirement of an open education program is more effective preparation of the teachers. While some programs in teacher preparation for open education are currently offered, the relative values of such programs will apparently depend upon their success in rendering explicit the goals and objectives of open education and they complement and differ from those of traditional education. Most models of open education seem to place more emphasis upon methodology than upon objectives, whereas it would seem more appropriate for methodology to follow objectives. Even the few available instruments to assess educational openness seem more materials- and method-oriented than goal-oriented. Much in need is a clearer statement of the behavioral objectives of open education.

2. The transition from a traditional program to a more open program may be difficult for many students, requiring new levels of self-discipline and perhaps of self-confidence as well. New habits of study and of pacing will be needed. In return for his greater freedom, the student will have to learn to work efficiently in a sometimes chaotic environment, and he will have to learn to depend more upon himself and less upon his teachers for control and guidance. These skills and habits are not emphasized in traditional education; hence the open classroom teacher well may have to teach things largely unrelated to the academic curriculum. And it well may be that development of the skills and habits necessary to survival and continued academic progress are the distinctive characteristics and most important products of open education. Further, it seems

likely that the more prolonged the student's experience in traditional education, the more important and the more difficult the acquisition of these skills and habits will be.

3. Little attention seems to have been paid to the potential influences of school personnel not directly involved in open education projects. Probably the assistance, support, and encouragement of administrators and other school personnel are more critical to the effectiveness of an innovative project than is commonly realized. Educators, like other humans, are frequently reluctant to accept innovation, particularly if they are not direct participants. Similarly, the innovation participants can easily feel threatened by a possibility of failure or isolated from accepted practices, and in such circumstances they may regress to behaviors inappropriate to the new situation. Because of its self-proclaimed variance from traditional education, open education must recognize that teachers, too, require guidance and leadership.

The open education program evaluated in this study will continue, and evaluation of it will also continue. The research model incorporated into the design of this study has proved its value, and as the concepts of open education are more closely realized in the program, this model will continue to be employed in evaluation of the merits of open education.

References

Barth, R. S., & Rathbone, C. H. A bibliography of open education. Newton, Mass.: Educational Development Center, 1971.

Bell, H. M. Manual for The School Inventory. Palo Alto, Cal.: Consulting Psychologists Press, 1963.

Burnham, B. Open education: Some research answers to basic questions. Orbit, 1971, 2, 22-24.

Carbonari, J. P. Report of an evaluation study of an open-concept school. Educators Report and Fact Sheet, 1971, 8, 5.

Educational Testing Service. Analysis of an approach to open education. Princeton, N. J.: Author, 1970.

Elementary Science Study. The ESS Reader. Boston: Educational Development Center, 1970.

Featherstone, J. Experiments in learning. The New Republic, 1968, 159 (24), 23-25.

Featherstone, J. Open schools I: The British and U.S. The New Republic, 1971, 165 (11), 20-25. (a)

Featherstone, J. Open schools II: Tempering a fad. The New Republic, 1971, 165 (13), 17-21. (b)

Featherstone, J. Schools where children learn. New York: Liveright, 1971. (c)

Flanders, N. A. Interaction analysis in the classroom. Ann Arbor, Mich.: University of Michigan Press, 1966.

Gough, H. G. Manual for the California Psychological Inventory. Palo Alto, Cal.: Consulting Psychologists Press, 1969.

Hawkins, D. Messing about in science. Science & Children, 1965, 2 (5), 5-9.

Hawkins, D. The informed vision: An essay on science education. In J. Kagan (Ed.), Creativity and learning. Boston: Houghton-Mifflin, 1967.

Hayes, W. L. Statistics. New York: Holt, Rinehart and Winston, 1963.

Howard, L. M. The developmental classroom. Boston: Dept. of Title I Programs, Boston Public Schools, 1968.

Kennedy, V. J., & Say, M. W. Comparison of the effects of open-area vs. closed-area classrooms on the cognitive gains of students. Educators Report and Fact Sheet, 1971, 8, 4.

Killough, C. K. An analysis of the longitudinal effects that a nongraded elementary program conducted in an open space school had on cognitive achievement of pupils. Houston: Bureau of Education Research & Services, University of Houston, 1971.

Plowden, Lady Bridget et al. Children and their primary schools: A report of the Central Advisory Council for Education. Vol. I. London: Her Majesty's Stationery Office, 1966.

Rathbone, C. H. (Ed.) Open education: The informal classroom. New York: Citation, 1971.

Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 1966, 80 (1, Whole No. 609).

Sackett, J. W. A comparison of self-concept and achievement of sixth grade students in an open-space school, self-contained school, and departmentalized school. Unpublished doctoral dissertation, University of Iowa, 1971. (Dissertation Abstracts, 32, 2372-A)

Silberman, C. E. Crisis in the classroom. New York: Random House, 1970.

Traub, R., Weiss, J., Fisher, C., & Musella, D. Closure on openness in education. Paper presented at the meeting of the American Educational Research Association, New Orleans, February, 1973.

Walberg, H. J., & Thomas, S. C. Characteristics of open education: Toward an operational redefinition. Newton, Mass.: Educational Development Center, 1971.

Walberg, H. J., & Thomas, S. C. Open education: An operational definition and validation in Great Britain and United States. American Educational Research

Journal, 1972, 9, 197-208.

Warner, J. B. A comparison of students' and teachers' performance in an open area facility and in self-contained classrooms. Unpublished doctoral dissertation, University of Houston, 1971. (Dissertation Abstracts, 31, 3851-A)

Wilson, F. S., Langevin, R., & Stuckey, T. Are pupils in the open plan school different? The Journal of Educational Research, 1972, 66, 115-118.

Table 2.
Intercorrelations of OSCAR Teachers' Mean Ratings of Students

	Math.	Social Studies	Science
Language	.427	.450	.395
Mathematics		.496	.861
Social Studies			.516

Table 1.
Patterns of significant differences (univariate F tests)
between various subject groupings.

Variable	Success vs. Failure N = 39/78	OSCAR Success vs. OSCAR Failure N = 50/45	Control Success vs. Control Failure N = 39/33	OSCAR Success vs. Control Success N = 50/39	OSCAR Failure vs. Control Failure N = 45/33
Paragraph meaning	30.12***[a]	15.10***[a]	15.08***[a]		
Spelling	44.99***[a]	32.65***[a]	13.40***[a]		
Language	18.95***[a]	10.30**[a]	8.71**[a]		
Arithmetic computation	21.47***[a]	4.10*[a]	31.95***[a]		
Arithmetic concepts	34.43***[a]	15.59***[a]	18.79***[a]		
Arithmetic application	17.41***[a]	6.14*[a]	12.32***[a]		
Social studies	39.56***[a]	24.77***[a]	15.77***[a]		
Science	16.52***[a]	7.70**[a]	8.92**[a]		
Sense of well-being				4.39*[b]	
Responsibility	23.03***[a]	9.25**[a]	14.63***[a]		
Communality		7.28**[a]			5.21*[b]
Self-concept				5.95*[b]	
Locus of control		4.21*[c]			

a=first group higher
b= second group higher
c=first group lower, here meaning more internalized

*p < .05
**p < .01
***p < .001